Central European Forum for
Migration and Population Research

Środkowoeuropejskie Forum
Badań Migracyjnych i Ludnościowych

Bayesian Model Selection in Forecasting International Migration: Simple Time Series Models and Their Extensions

Jakub Bijak

Central European Forum for Migration and Population Research

Joint Eurostat / UN ECE Work Session on Demographic Projections

Bucharest, 10–12 October 2007, Session 7: Specific Projection Issues

Project financed from the research grant of the Foundation for Population, Migration and Environment (BMU-PME) from Zurich

Plan of the presentation

1. Bayesian forecasting: general remarks
2. Bayesian model selection: introduction
3. Empirical application: Polish–German migration forecasts for 2005–2015
    - Simple stochastic processes – ARMA(1,1) sub-models
    - Models with changes in conditional variance
    - Forecasting by analogy to other migration flows
4. Robustness of forecasts against selected changes in the prior distributions
5. Concluding remarks

1. Bayesian forecasting: general remarks

Predictive distribution of forecasted values $\mathbf{x}^F$ given data $\mathbf{x}$, weighted by posterior distributions of parameters θ, equals:

$$p(\mathbf{x}^F|\mathbf{x}) = \int_\Theta p(\mathbf{x}^F|\mathbf{x}, \theta) \cdot p(\theta|\mathbf{x})\, d\theta,$$

where, by the Bayes' Theorem (1763):

$$p(\theta|\mathbf{x}) = p(\mathbf{x}|\theta) \cdot p(\theta) / p(\mathbf{x}); \quad p(\mathbf{x}) = \int_\Theta p(\mathbf{x}|\theta) \cdot p(\theta)\, d\theta$$

Posterior $p(\theta|\mathbf{x})$ is proportional to data likelihood $p(\mathbf{x}|\theta)$ and prior distribution $p(\theta)$, the latter denoting subjective expert knowledge (judgement) on the model parameters

Probability in the Bayesian statistics is also a subjective concept, interpreted as a measure of belief

2. Bayesian model selection: introduction

Theory: Uncertainty of models

- Let M_1, …, M_m be mutually exclusive (not nested) models adding up to a finite space of possible models, **M**
- Let $p(M_1)$, …, $p(M_\mathrm{m})$ be the models' prior probabilities, e.g.:
  - Flat prior (equal probabilities): $p(M_1) = \ldots = p(M_\mathrm{m})$
  - "Occam's razor" prior, favouring simpler models with smaller numbers of parameters, l_i: $p(M_i) \propto 2^{-l_i}$
- For forecasting, a model with highest posterior probability is selected on the basis of the Bayes' Theorem:

$$p(M_i|\mathbf{x}) = p(M_i)\cdot p(\mathbf{x}|M_i)\,/\,\Sigma_{k\in \mathbf{M}}\,\{p(M_k)\cdot p(\mathbf{x}|M_k)\}$$

[Hoeting et al., 1999; Osiewalski, 2001]

2. Bayesian model selection: introduction

Practice: Numerical computations

- Model Choice via Markov Chain Monte Carlo [Carlin and Chib, 1995] implemented in WinBUGS 1.4 software [Spiegelhalter et al., 2003]
- Method: iterative sampling from full conditional distributions for model-specific parameters θ_i and the model index μ:

$$\begin{cases} p(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}_{j\neq i}, \mu, \mathbf{x}) \propto \begin{cases} p(\mathbf{x} \mid \boldsymbol{\theta}_i, \mu = i) \cdot p(\boldsymbol{\theta}_i \mid \mu = i) & \text{for } \mu = i \\ p(\boldsymbol{\theta}_i \mid \mu \neq i) & \text{for } \mu \neq i \end{cases} \\ p(\mu = i \mid \boldsymbol{\theta}, \mathbf{x}) = \dfrac{p(\mathbf{x} \mid \boldsymbol{\theta}_i, \mu = i) \cdot p(M_i) \cdot \prod_{j\in\mathbf{M}} p(\boldsymbol{\theta}_j \mid \mu = i)}{\sum_{k\in\mathbf{M}} [p(\mathbf{x} \mid \boldsymbol{\theta}_k, \mu = k) \cdot p(M_k) \cdot \prod_{j\in\mathbf{M}} p(\boldsymbol{\theta}_j \mid \mu = k)]}. \end{cases}$$

- Parameters θ_i are sampled either from full conditionals if μ = *i*, or from linking densities ("pseudo-priors") otherwise
- Iterations before reaching convergence are discarded

2. Bayesian model selection: introduction

Rationale for migration forecasting applications

- Features of the Bayesian approach:
  - The stochastic character ensures the formality of inference, with key focus on the uncertainty issue
  - The *a priori* expert judgement is allowed, which can supplement small-sample information (important for many time series of European migration) [cf. Willekens, 1994]
- Formal model selection techniques:
  - One way of assessing the uncertainty of model specification
  - When used with appropriate priors (e.g., "Occam's razor"), can answer the question on simplicity *versus* complexity in the population forecasting models [cf. Ahlburg, 1995; Smith, 1997]

3. Forecasts of Polish–German migration

Aim

To forecast long-term migration between Germany and Poland for 2005–2015 in different modelling frameworks

Data

- Forecasted variable – logarithms of emigration rates per 1,000 population of the sending country:

$$m_t = \ln(Mig_t / Pop_t * 1{,}000)$$

- Data series for (1985–)1991–2004
- Data sources: population stocks – Eurostat
  migration – Destatis (German data)
- Polish population stocks include post-census adjustment

3. Forecasts of Polish–German migration

a) Simple stochastic processes – ARMA(1,1) sub-models

- M_1: $m_t = c + \varepsilon_t$ [oscillations around a constant]
- M_2: $m_t = c + m_{t-1} + \varepsilon_t$ [random walk with drift]
- M_3: $m_t = c + \phi\, m_{t-1} + \varepsilon_t$; $\phi \notin \{0, 1\}$ [AR(1) process]
- M_4: $m_t = c - \theta\, \varepsilon_{t-1} + \varepsilon_t$; $\theta \neq 0$ [MA(1) process]
- M_5: $m_t = c + \phi\, m_{t-1} - \theta\, \varepsilon_{t-1} + \varepsilon_t$; $\phi,\ \theta \neq 0$ [ARMA(1,1)]

Random term: $\varepsilon_t \sim$ iiN$(0, \sigma^2)$ **Sample:** 1991–2004 (N=14)

Priors: constants $c \sim$ N$(0, 100^2)$ diffuse (hardly informative)

$\phi,\ \theta \sim$ N$(0.5, 1^2)$: processes likely stationary / time-reversible

Low precision: $\tau_{\mathrm{PL}\rightarrow\mathrm{DE}} = \sigma^{-2} \sim \Gamma(0.25, 0.25)$; $\tau_{\mathrm{DE}\rightarrow\mathrm{PL}} \sim \Gamma(4, 0.4)$

3. Forecasts of Polish–German migration

b) AR(1) extensions: non-constant conditional variance

General model: $m_t = c + \phi\, m_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim \mathrm{N}(0, \boldsymbol{\sigma_t^2})$

- M_5: $\sigma_t^2 = \sigma^2$ [reference model, constant variance]
- M_6: $\sigma_t^2 = k + \alpha \cdot \varepsilon_{t-1}^2$ [AR(1)–ARCH(1) process]
- M_7: $\sigma_t^2 = k + \alpha \cdot \varepsilon_{t-1}^2 + \beta \cdot \sigma_{t-1}^2$ [AR(1)–GARCH(1,1)]
- M_8: $\ln(\sigma_t^2) = k + \gamma \cdot \ln(\sigma_{t-1}^2) + \zeta_t$ [simple stochastic volatility, SV]

Deterministic (M_6–M_7) vs. stochastic (M_8) changes in variance

2nd random term: $\zeta_t \sim \mathrm{iiN}(0, \rho^2)$ **Sample:** 1985–2004 (N=20)

Priors: c, ϕ as before, other concentrated for computational reasons: $\alpha, \beta, \gamma \sim \Gamma(10, 20)$; $k \sim \Gamma(1, 0.1)$; $1/\rho^2 \sim \Gamma(10, 1)$

3. Forecasts of Polish–German migration

c) Models with analogy to Iberian migration flows

Idea: to capture institutional changes, e.g., post-accession opening of the Western EU labour markets [Kupiszewski, 1998]

- M_{10}: $m_t = c + \varepsilon_t$ [reference model, no analogy]
- M_{11}: $m_t = c + a \cdot m^{PT}{}_{t-18} + b \cdot \mathbf{1}_{t=2002} + \varepsilon_t$ [Portugal]
- M_{12}: $m_t = c + a \cdot m^{ES}{}_{t-18} + \varepsilon_t$ [Spain]
- M_{13}: $m_t = c + a \cdot m^{IB}{}_{t-18} + b \cdot \mathbf{1}_{t=2002} + \varepsilon_t$ [both countries]

Rationale: timing of EU accession, system transformation

Random term: $\varepsilon_t \sim$ AR(1) **Sample:** 1992–2004 (N=13)

Priors: c, ϕ, τ as before, $a \sim$ N(0.5, 1^2) – a positive analogy

3. Forecasts of Polish–German migration

Bayesian model selection for three proposed classes M

Model posteriors $p(M_i|\mathbf{x})$ under "Occam's razor" priors, $p(M_i) \propto 2^{-l_i}$

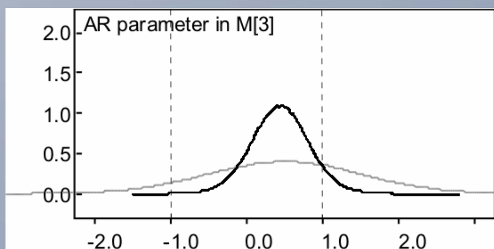
Migration flow	Subclasses of ARMA(1,1)					Extensions of variance				Models with analogy			
	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	M_{13}
Poland → Germany	**0.42**	0.29	0.14	0.11	0.03	0.12	0.10	0.05	**0.74**	**0.69**	0.07	0.14	0.10
Germany → Poland	0.23	**0.49**	0.16	0.08	0.03	**0.71**	0.02	0.00	0.27	**0.88**	0.01	0.09	0.02
Model priors $p(M_i)$	0.31	0.31	0.15	0.15	0.08	0.50	0.25	0.13	0.13	0.40	0.20	0.20	0.20

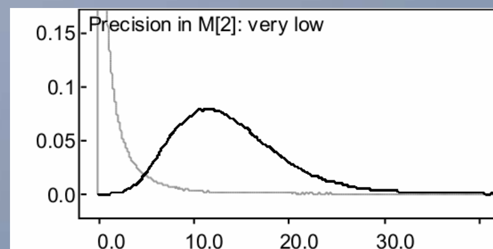
- Simple random models (oscillations / random walks)
- Either constant or stochastic conditional variance
- No linear analogies supported by the data

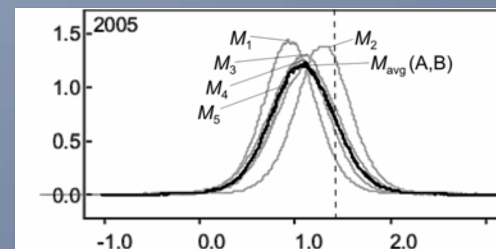
3. Forecasts of Polish–German migration

Overview of selected empirical results

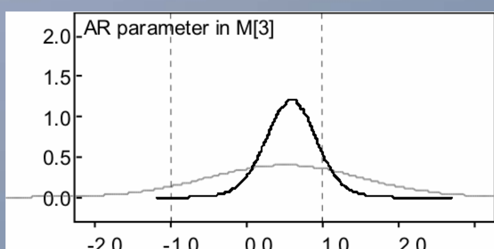
Migration from Poland to Germany: selected distributions, forecast for 2005

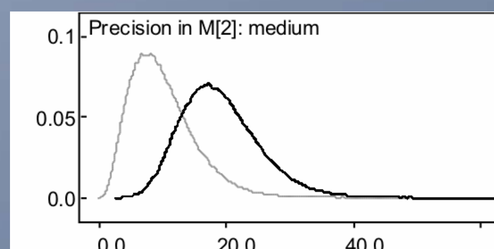


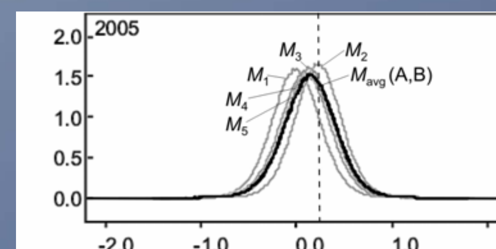




Migration from Germany to Poland : selected distributions, forecast for 2005







Grey lines – distributions *a priori*, black – *a posteriori*.

Highest *ex post* errors for M_1, lowest – for M_2

3. Forecasts of Polish–German migration

Exp(m_t) forecasts from the selected models, 2005–2015

Model	Forecasted exp(m_{2005})			Forecasted exp(m_{2010})			Forecasted exp(m_{2015})		
	10%	Median	90%	10%	Median	90%	10%	Median	90%
Poland → Germany: exp(m_{2004}) = **3.65;** exp(m_{2005}) = **4.17**									
M_1: oscillation	**1.77**	**2.57**	**3.72**	**1.78**	**2.57**	**3.71**	**1.78**	**2.57**	**3.72**
M_9: AR(1)–SV	**2.64**	**3.42**	**4.41**	**1.74**	**3.31**	**7.18**	**1.60**	**3.28**	**8.40**
M_{10}: no analogy	**1.96**	**2.99**	**4.41**	**1.71**	**2.62**	**4.04**	**1.71**	**2.63**	**4.05**
Germany → Poland: exp(m_{2004}) = **1.27 ;** exp(m_{2005}) = **1.28**									
M_2: RWD	**0.91**	**1.25**	**1.71**	**0.48**	**1.18**	**2.90**	**0.28**	**1.12**	**4.36**
M_6: AR(1)	**0.90**	**1.26**	**1.76**	**0.69**	**1.24**	**2.36**	**0.63**	**1.24**	**2.71**
M_9: AR(1)–SV	*0.94*	*1.25*	*1.67*	*0.72*	*1.21*	*2.24*	*0.68*	*1.21*	*2.62*
M_{10}: no analogy	**0.80**	**1.12**	**1.55**	**0.69**	**1.01**	**1.48**	**0.69**	**1.01**	**1.48**

- Median trajectories plausible, indicate stabilisation
- Limits of 80-percent predictive intervals reasonable, except for the (likely) non-stationary models (RWD/AR)

4. Robustness against changes in priors

a) Uniform model priors $p(M_i)$ instead of the "Occam's razor"

- Results for ARMA(1,1) sub-models: the same ones selected, although with different posterior probabilities $p(M_i|\mathbf{x})$

b) Alternative prior distributions for θ, used as a reference: the non-informative ones [Jeffreys, 1961]

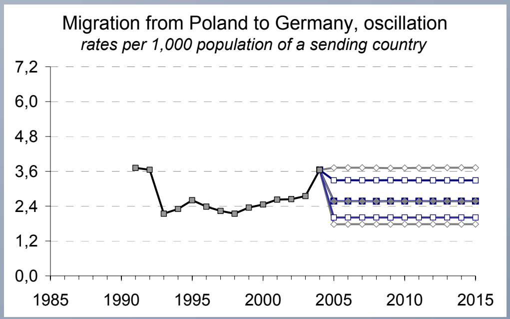
- In practical applications, "hardly informative" priors can be used for computational convenience: [Congdon, 2003]

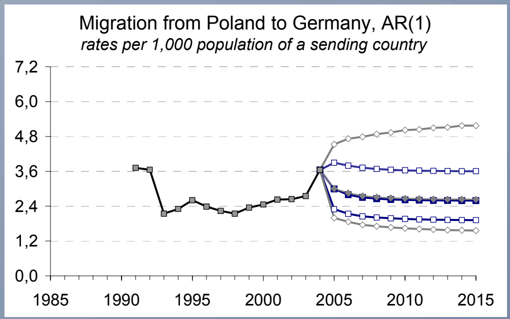
  For structural parameters: N(0, D^2), where D is a big number

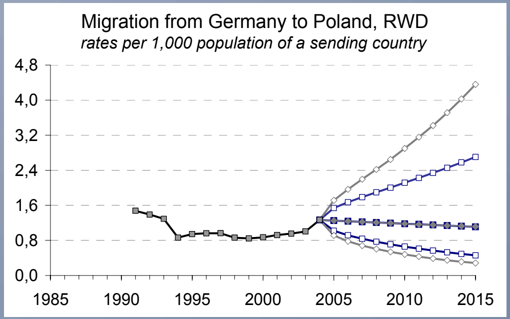
  For precision $\tau = \sigma^{-2}$: $\Gamma(a, a)$, with small parameters a
- Under D = 100 and a = 0.001, the obtained forecasts for oscillations, random walks and AR(1) differ from the "informative" ones with respect to uncertainty estimates

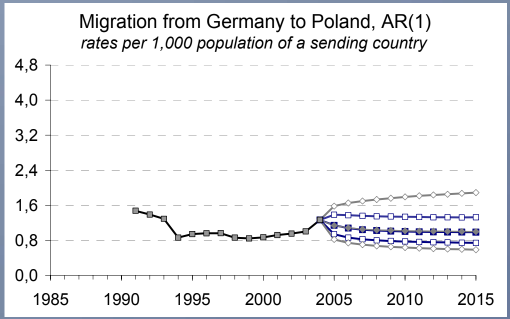
4. Robustness against changes in priors

Results: informative *vs.* hardly informative priors on precision









Without *a priori* assumptions on low precision τ, in many cases the 80-percent intervals are narrower than within-sample variability

5. Concluding remarks

- Bayesian model selection techniques allow for identifying models with the highest data support and for assessing uncertainty on various levels, including model specification
- Empirical results: simple, unstructured models preferred
- Selection of oscillations, random walks and stochastic volatility models confirms a hardly predictable character of both migration rates and their uncertainty measures
- Given the shortness of data series, the results are not robust against changes in priors (especially for precision)
- However, without assuming low precision *a priori*, the predictive intervals would be in many cases too narrow as for such an uncertain phenomenon as migration

Thank you for your attention!

Paper first presented during the conference “Dynamic Econometric Models” in Toruń, Poland, 4–6 September 2007. Credits for several underlying ideas, as well as for helpful remarks and discussions go to Prof. Jacek Osiewalski.